

# EMSContExt: EMS Protocol-driven Concept Extraction for Cognitive Assistance in Emergency Response

Sarah M. Preum<sup>1</sup>, Sile Shu<sup>2</sup>, Homa Alemzadeh<sup>2</sup>, John A. Stankovic<sup>1</sup>

<sup>1</sup>Department of Computer Science, <sup>2</sup>Department of Electrical and Computer Engineering  
University of Virginia, Charlottesville, Virginia, 22904  
{preum, ss5de, ha4d, jas9f}@virginia.edu

## Abstract

This paper presents a technique for automated curation of a domain-specific knowledge base or lexicon for resource-constrained domains, such as Emergency Medical Services (EMS) and its application to real-time concept extraction and cognitive assistance in emergency response. The EMS responders often verbalize critical information describing the situations at an incident scene, including patients' physical condition and medical history. Automated extraction of EMS protocol-specific concepts from responders' speech data can facilitate cognitive support through the selection and execution of the proper EMS protocols for patient treatment. Although this task is similar to the traditional NLP task of concept extraction, the underlying application domain poses major challenges, including low training resources availability (e.g., no existing EMS ontology, lexicon, or annotated EMS corpus) and domain mismatch. Hence, we develop EMSContExt, a weakly-supervised concept extraction approach for EMS concepts. It utilizes different knowledge bases and a semantic concept model based on a corpus of over 9400 EMS narratives for lexicon expansion. The expanded EMS lexicon is then used to automatically extract critical EMS protocol-specific concepts from real-time EMS speech narratives. Our experimental results show that EMSContExt achieves 0.85 recall and 0.82 F1-score for EMS concept extraction and significantly outperforms MetaMap, a state-of-the-art medical concept extraction tool. We also demonstrate the application of EMSContExt to EMS protocol selection and execution and real-time recommendation of protocol-specific interventions to the EMS responders. Here, EMSContExt outperforms MetaMap with a 6% increase and six times speedup in weighted recall and execution time, respectively.

## 1 Introduction

Based on a 2014 study, there are over 250,000 emergency medical service (EMS) providers in the U.S. who provide care for over 30 million emergency incidents annually (Duong et al. 2018). Upon arrival at an incident scene, the

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

first responders, including EMS providers, assess the situation and provide primary medical care to the victims before transferring them to hospitals. They collect a substantial amount of data from the scene (e.g., victim's present conditions, past medical history) and combine it with the knowledge of established EMS protocols<sup>1</sup> to take effective response actions. The tasks of collecting and interpreting this data and recalling the correct set of protocols to perform from hundreds of overlapping EMS protocols in such a high-stress situation are error-prone and cognitively demanding.

Hence, we envision an intelligent cognitive assistant system (Preum et al. 2019; 2018) that can reduce the first responders' cognitive burden by (i) real-time collection and interpretation of speech and other data from the incident scene and (ii) providing context-specific feedback on the safe response actions to take. This paper presents *EMSContExt*, a concept extraction system that is a critical component in the proposed cognitive assistant and identifies relevant terms from the textual transcriptions of the spoken language at the scene and maps them into EMS protocol-specific concepts. It is essential for the proposed cognitive assistant to identify relevant protocols that should be executed in each context. Here we refer to the joint tasks of concept identification and mapping as concept **extraction** or annotation.

CC: chest back and R arm pain ...  
A: pt A+Ox4 ...  
A: patent B: rate and quality good no sob reported C:b skin  
N/D/W  
pulses = reg rapid at 110 b/p as noted ...  
E: No visible trauma bleeding or resp distress noted

**Text Box 1:** An excerpt from an EMS narrative. The underlined concepts are critical for cardiac-suspected chest pain protocol.

Referring to the EMS narrative excerpt shown in Text Box 1, the goal of *EMSContExt* is to automatically identify the protocol-specific terms, such as skin N/D/W and mapping it to the standard concept of skin is normal,

<sup>1</sup>EMS protocols are standard procedures designed by the national, state, and local medical and EMS authorities. The EMS providers are trained to operate according to these protocols. Sample EMS protocols can be found in <http://www.emsprotocols.org>.

dry, and warm. This concept extraction problem involves the following challenges that limit the applicability of existing medical concept identification and normalization solutions (Luo et al. 2018), (Mujjiga et al. 2019) to the EMS domain. Firstly, pre-hospital EMS data often contains terms, abbreviations, and short-hands that are unique to the EMS domain, e.g., in Text Box 1 A+O×4 is a unique EMS short-hand that refers to "patient is alert and oriented of surrounding person, place, time, and situation". Although the EMS domain significantly overlaps with the medical domain, the former has many unique terminologies (e.g., A+O×4, N/D/W) that are critical for EMS protocol execution but do not appear in a medical or generic text. Thus existing solutions for concept extraction from the medical text can not be directly applied here due to **domain mismatch**. Secondly, the EMS domain has very **low training resources** as there is no protocol-specific ontology, knowledge-base, or corpus of EMS transcripts, let alone any EMS corpus with annotation of protocol-specific concepts. To make matters worse, the manual annotation of EMS concepts is prohibitively expensive in terms of both time and intellectual effort. Thus, it limits the applicability of existing concept extraction solutions that require reasonable amounts of data.

To address these challenges, we develop *EMSContExt*, a weakly-supervised system for extracting EMS protocol-specific concepts from responders' conversations at the scene. In order to identify all the variations of the instances of a concept mentioned in the spoken language, we develop a knowledge-integrated, data-driven weakly-supervised lexicon expansion approach for EMS concepts. Specifically, *EMSContExt* utilizes multiple knowledge bases and distributional semantic concept models developed using unstructured EMS scene descriptions to curate a lexicon of critical EMS concepts automatically. The lexicon of EMS concepts (i.e., the original concepts and their frequent lexical variations) are then used to extract protocol-specific safety-critical concepts from the EMS conversations in real-time to facilitate cognitive assistance. The main contributions of the paper are as follows:

(i) We develop a novel, weakly-supervised approach to automatically curate domain-specific knowledge bases or lexicons for resource-constraint, safety-critical domains, such as EMS. Our approach requires one-time minimum supervision from domain experts and utilizes standard lexical and medical knowledge bases and word embedding models trained on a domain-specific corpus of over 9400 EMS narrations. The curated EMS lexicon, word embedding models, and textual EMS corpus developed as part of this work will be shared publicly to accelerate further research on intelligent applications for emergency response.

(ii) *EMSContExt* achieves 0.85 recall and 0.82 F1-score on average for protocol-specific concept extraction and outperforms the state-of-the-art medical concept extraction tool, MetaMap, with a three times increase in F1-score and 21% increase in recall.

(iii) We evaluate the applicability of *EMSContExt* for accurate selection and execution of EMS protocols and recommending EMS protocol-specific interventions by the cognitive assistant. Using *EMSContExt*, we achieve, respec-

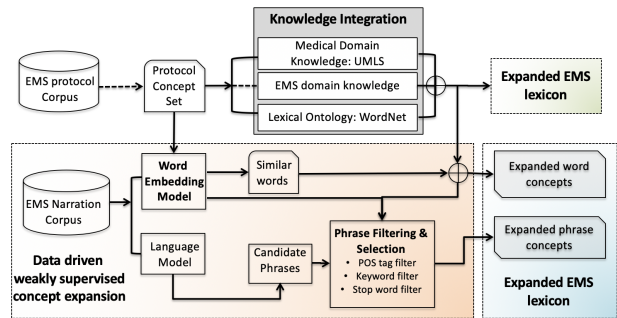


Figure 1: Overview of lexicon expansion in *EMSContExt*.

tively, a 4% increase, 6% increase, and six times speedup in weighted F1-score, weighted recall, and execution time for recommending EMS interventions compared to MetaMap.

## 2 Approach

The lexicon expansion pipeline of *EMSContExt* is shown in Figure 1. It starts with collecting a set of seed protocol-specific concepts ( $S_{PC}$ ) from a regional EMS protocol corpus. Each concept  $C_i$  in  $S_{PC}$  represents a condition related to the patient or scene that should be checked by the responders before performing an intervention at the scene. Thus our concept extraction approach is **protocol-driven** to facilitate cognitive assistance through selection and execution of most relevant protocols (Shu et al. 2019). The seed lexicon of EMS concepts is then expanded to a larger dictionary of terms by utilizing multiple knowledge bases that capture the lexical variations of the concepts (grey box in Figure 1). To further improve the lexicon expansion process, a novel weakly-supervised data-driven method is developed (the orange box in Figure 1). The output of this knowledge-integrated, data-driven lexicon expansion approach is a lexicon of EMS terms (green and blue boxes in Figure 1), which is then used to automatically extract the occurrences of the lexical variations of EMS concepts from EMS narrations.

### Knowledge-integrated Lexicon Expansion

Our empirical observations from a review of EMS narratives suggest that some variations of an EMS concept are linguistic while others are stemmed from the medical and EMS domains. For example, 'respiration rate' is synonymous to *breathing rate* (i.e., a linguistic variation) and can also be expressed as *rr* as a short-hand (i.e., a domain-based abbreviation). Hence, *EMSContExt* integrates multiple knowledge bases as described next (shown in the grey box in Figure 1).

**Integrating lexical knowledge:** WordNet (Miller 1995) is one of the most popular lexical ontologies for the English language. It groups nouns, verbs, adjectives, and adverbs into cognitive synonyms or synsets.

For each concept  $C_i$  of set  $S_{PC}$ , corresponding synsets, including synonyms, antonyms, hypernyms, and hyponyms are automatically expanded using the WordNet API. The synsets are annotated to identify relevant word senses.

**Integrating medical knowledge:** Since the EMS domain overlaps with the medical domain, medical knowledge is integrated using the Unified Medical Language System (UMLS). Its the state-of-the-art unified platform of key terminology, ontologies, and other resources used in medical settings. For each concept  $C_i$  of the set  $S_{PC}$ , a set of relevant medical concepts are expanded using the UMLS API.

**Integrating EMS knowledge:** A knowledge base of unique EMS abbreviations and short-hand expressions is developed to integrate EMS domain knowledge, e.g., N/D/W or A+Ox4. Two EMS providers compiled a dictionary of 370 such terms and their actual meanings. Additional abbreviations were integrated from the national EMS guidelines<sup>2</sup>.

These three knowledge bases are integrated to create a novel dictionary of protocol-specific EMS concepts that can be used for lexicon expansion. However, since this approach captures only a subset of terms similar to a given concept, we further incorporated data-driven lexicon expansion.

### Data-driven Lexicon Expansion

Distributional semantic models have been increasingly used in the NLP research to capture words that have semantic similarity (Mikolov et al. 2013; Pennington, Socher, and Manning 2014). But the performance of generic word embedding models drastically drops when applied to a domain-specific text, especially for a low-resource, constrained vocabulary domain like EMS. *EMSContExt* addresses this challenge of domain mismatch by using a domain-specific distributional semantic model.

**Distributional semantic model:** The EMS corpus is pre-processed to replace low-frequency words and numbers with special tokens UNK and NUM, respectively. Next, the corpus is tokenized into sentences and words. Then, a word2vec model is trained on this EMS corpus<sup>3</sup>. The hyper-parameters of the word2vec model are tuned using a validation set. This model is used to query relevant contextual terms that are similar to a given concept in  $S_{PC}$ . Some variations of a concept are unigrams, and some are n-grams or phrases. For example, the concept `trauma` can be mapped to `injury` or `bruising` and `abrasions`.

**Word level concept expansion:** The word embedding model developed in the previous step is queried with a concept  $C_i$  to return the top  $k$  words similar to  $C_i$  to expand the word level concepts. The returned set of concepts is filtered to drop: (i) any stop words using a domain-specific stop word list<sup>4</sup> and (ii) any antonyms using the lexical EMS dictionary. For example, for the concept `hypertension`, if the set of terms expanded by the word embedding model contains its antonym (i.e., `hypotension`), the antonym should be discarded. The filtered set of terms is considered the expanded word-level lexicon of the concept  $C_i$ .

**Phrase level concept expansion:** To expand phrase level variations of a concept, we develop bigram, trigram, and

quadgram language models from the training data. Language models with higher n-grams (i.e.,  $n > 4$ ) can be incorporated, if necessary. These models are used to generate potential candidate phrases similar to the original concepts. Next, these potential phrases are scored and ranked according to their frequency and point-wise mutual information gain (Church and Hanks 1990). The resulting set of candidate phrases are filtered using the following three filters to find phrases similar to a given concept efficiently:

(i) **Stop word filter:** Empirically, the candidate phrases containing a stop word are much less informative than the candidate phrases without any stop word. So, the candidate phrases generated from the language models are filtered to drop any phrase that contains at least one stop word using the domain-specific stop word list.

(ii) **Keyword filter:** A list of keywords ( $KW_i$ ) for each concept  $C_i$  in  $S_{PC}$  is automatically compiled from the EMS dictionary developed using knowledge-integrated lexicon expansion and the expanded word-terms from the word embedding model that represents the set of terms similar to  $C_i$ . The potential candidate phrase set filtered from the previous step is further filtered to drop the phrases that do not contain any of the keywords in  $KW_i$ . Also, the list of antonyms is used to filter the phrases that contain antonym of  $C_i$ .

(iii) **Parts-of-speech tag filter:** A parts-of-speech (POS) tag filter is used to eliminate less relevant candidate phrases. This filter parses through the POS tags of a candidate phrase and drops the phrases that contain specific patterns of POS tags. For example, in the text `patient reported severe pain`, the POS tag filter drops the phrase consisting of the first two words (noun followed by a verb) since it cannot be mapped to a protocol-specific concept. However, the phrase consisting of the last two words (adjective followed by a noun) is similar to the concept of `pain`.

*EMSContExt* uses average cosine similarity to calculate the similarity between the distributional representations of the given concept and each filtered candidate phrase. The candidate phrases are ranked according to their similarity score, and the top  $k$  candidate phrases with the highest similarity scores are returned as the expanded phrases related to a given original concept.

The expanded EMS lexicon resulting from the knowledge-integrated, data-driven pipeline (the green and blue boxes in Figure 1) is then used to automatically extract protocol-specific concepts (the original concepts and their lexical variations) from EMS narratives using string matching. Additional details about the implementation would be made available in a technical report. To summarize, *EMSContExt* addresses the challenges of **domain mismatch** and **low training resources** using weak supervision that does not require any training data and utilizes minimal, one-time supervision from domain experts for expanding the EMS lexicon.

## 3 Experimental Setting

**Heterogeneous textual corpora:** *EMSContExt* utilizes multiple EMS and Medical textual corpora for training distributional semantic model as described in Section 2. They vary in modality, content, and structure.

<sup>2</sup><https://tinyurl.com/yxu25zpp>

<sup>3</sup>Any other word embedding model can be used too.

<sup>4</sup>The domain-specific stop word list includes standard English stop words and additional stop words specific to the EMS domain, e.g., `pt.` (a short-hand of `patient`).

(i) EMS corpus: The EMS corpus, RAA, contains 8000 real de-identified EMS records from a regional ambulance agency. Each record consists of the following attributes: call type, chief complaint, impression, vitals, interventions, and narratives. The first five fields are structured or semi-structured, and the narrative field is unstructured. The narrative contains additional critical information, e.g., dispatch, history, assessment. The RAA dataset contains long, distilled, and detailed EMS narratives.

(ii) Medical corpora: For generating word embedding models, we have also considered the medical text, since they significantly overlap with EMS text. Two benchmark medical textual datasets are used: I2B2 challenge dataset (Uzuner et al. 2011) and MIMIC III dataset (Johnson et al. 2016). For both of these datasets, we used the only data that we got access to upon request. The I2B2 challenge dataset contains a total of 5733 de-identified electronic health records collected over ten years. The MIMIC III dataset we used consists of 2434 de-identified nursing notes.

The following word embedding models are generated using the above mentioned datasets as training data: (i) only EMS corpus, (ii) only medical corpora, and (iii) using union of EMS and medical corpora. However, we found that the word embedding model using only EMS corpus results in better performance for concept extraction. So, we have used that model for the experiments described in Section 4.

**EMS protocol corpus and Test dataset:** A set of 8 of the most common EMS protocols is selected from a regional EMS council for evaluation. This includes chest pain, seizure, respiratory distress, altered mental status, and opioid overdose. From this protocol set, 85 critical seed concepts were collected by two real responders as the core concepts. Each of these concepts represents a condition related to the patient or scene that should be checked by the responders for deciding on the most effective intervention to be performed at the scene. Eighty-four cases from the RAA data are annotated according to the concepts that trigger any of the eight selected protocols. Two EMS providers annotated the test data. The test data is disjoint from the training data used to train and validate the word embedding model.

## 4 Evaluation

In this section, the performance of *EMSContExt* is compared with multiple baselines and a state-of-the-art medical concept extraction tool. Besides, an end-to-end evaluation is provided to demonstrate the effectiveness of the concepts extracted by *EMSContExt* for EMS protocol execution and intervention suggestion by the cognitive assistant system.

### Comparing performance of *EMSContExt* with baseline and state-of-the-art

In this experiment, the performance of *EMSContExt* is compared with the following baseline and the state-of-the-art system for concept extraction. The whole knowledge-integrated weakly-supervised pipeline described in Section 2 is used in this experiment for concept expansion. Then the resulting expanded concept set is used for automatically ex-

tracting the occurrences of the EMS protocol-specific concepts from the unseen test data.

**Baseline,  $S_{PC}$ :** In this baseline,  $S_{PC}$  is used as keywords to identify protocol-specific concepts in test data. The goal of this baseline is to demonstrate the effect of using only the seed concept set ( $S_{PC}$ ) for concept extraction without any additional lexicon expansion.

**MetaMap:** The state-of-the-art medical concept extraction or annotation tool is MetaMap (Aronson and Lang 2010). It extracts medical concepts from text using an ensemble of knowledge bases and a linguistic rule base. It categorizes the extracted concepts into one of 133 semantic types, including symptoms, medications, activity, organs, and food. Since there is no concept extraction system for EMS, we use MetaMap as the next best available solution. The performance of MetaMap is evaluated with the following two settings based on the semantic types. **First**, running MetaMap in the default mode without filtering any semantic types. This baseline is referred to as **MetaMap**. **Second**, running MetaMap with a subset of relevant semantic types where the concepts that do not belong to any of the EMS relevant semantic types (e.g., plants, animals) are dropped. This is done to reduce the false positive rate, as MetaMap extracts concepts of several semantic types that are not relevant to the EMS domain. This baseline is referred to as **MetaMap-ST**.

The results of this experiment are presented in Table 1. The baseline using only  $S_{PC}$  results in the lowest recall among the four methods considered since it only contains the limited set of seed concepts. MetaMap-ST results in about 22% higher F1-score than MetaMap. But it results in about 12% lower recall than MetaMap, as a subset of the concepts extracted by MetaMap is filtered out in MetaMap-ST. *EMSContExt* outperforms both MetaMap-ST and MetaMap in terms of recall, precision, and F1 score. It yields about 32% and 65% higher recall than MetaMap and MetaMap-ST, respectively. It yields about three times higher F1 score than both MetaMap and MetaMap-ST. Thus, *EMSContExt* outperforms the state-of-the-art medical concept extraction tool. We also considered CLAMP (Soysal et al. 2017) as another baseline and found that MetaMap outperforms it in terms of both recall and F1-score.

Method	Recall	Precision	F1
$S_{PC}$	0.3874	0.9515	0.5506
MetaMap-ST	0.5155	0.20	0.2883
MetaMap	0.6435	0.1445	0.236
<i>EMSContExt</i>	<b>0.8538</b>	0.7855	<b>0.8182</b>

Table 1: *EMSContExt* outperforms both MetaMap and MetaMap-ST: it achieves about three times higher F1-score than both MetaMap and MetaMap-ST.

### Applying *EMSContExt* to Cognitive Assistance

We demonstrate the applicability of *EMSContExt* for protocol execution and intervention suggestions to provide real-time decision support to EMS providers. The EMS protocol selection and execution component in the proposed cognitive assistant models the EMS protocols used for de-

cision making in emergency response. It utilizes the outputs of the concept extraction stage to provide real-time cognitive support and feedback to first responders. For example, by extracting the patient’s status as unstable and EMS resources as no IV access from the scene narration, the cognitive assistant system can execute the Supraventricular Tachycardia protocol and suggest the synchronized cardioversion action to the responder. The details of this module can be found in (Shu et al. 2019) and are beyond the scope of the paper.

We formulate the problem of recommending EMS protocol-specific intervention suggestion as a multi-label classification problem where the input is the EMS narration, and the output is the set of potential interventions to perform at an emergency scene. For this experiment, the set of protocols selected is the same as the set of protocols of the RAA test dataset. A random set of 3657 EMS narrations are selected from the RAA data. We use *EMSContExt* to extract concepts from the input narration and use DEEPEN for negation detection. Then the output of the *EMSContExt* is fed as the input concepts to the protocol selection and execution component. The goal of this experiment is to measure the effectiveness of *EMSContExt* in extracting relevant concepts and improving the performance of the overall cognitive assistance pipeline. The performance of *EMSContExt* is compared with MetaMap. Since the classes are imbalanced (as some interventions are more common than others), we use weighted average recall, precision, and F1-score as the performance metrics due to class imbalance. The summary of the results is presented in Table 2.

Overall, *EMSContExt* yields about 6% higher weighted recall and 4% higher weighted F1 score in intervention suggestion than MetaMap and drastically reduces the execution time per iteration by about 6 times. The performance of *EMSContExt* affected by the following issue. The protocol selection requires semantic inference of the concepts and negation detection. Since *EMSContExt* does not perform negation detection, we have used DEEPEN (Mehrabi et al. 2015), a negation detection tool for detecting negation in medical text. On the other hand, in addition to concept extraction, MetaMap performs negation detection from medical text. So for MetaMap, we have used its built-in negation detection in this experiment. For negation detection in EMS text, the accuracy of DEEPEN is significantly lower than the accuracy of MetaMap. This has negatively affected the performance of *EMSContExt* for protocol execution. Hence, although *EMSContExt* yields about 21% higher recall for concept extraction than MetaMap, its increase in the recall has reduced for protocol execution and intervention suggestion due to the errors propagated from the negation detection tool. This issue will be addressed in the future by adapting DEEPEN for the EMS domain.

Although *EMSContExt* marginally improves the performance of protocol execution and intervention suggestion in terms of weighted recall and F1-score, it significantly reduces the **execution time**. Specifically, the execution time per iteration of *EMSContExt* is six times faster than MetaMap. It is crucial for an emergency response since the slightest delay in performing a critical intervention can have

Method	Weighted Recall	Weighted F1 score	Time (sec/case)
MetaMap	0.654	0.648	11.98
<i>EMSContExt</i>	<b>0.698</b>	<b>0.676</b>	<b>2.01</b>

Table 2: The performance of *EMSContExt* is compared with MetaMap for protocol execution and intervention suggestion. *EMSContExt* yields about 6% higher weighted recall and 4% higher weighted F1 score than MetaMap and reduces the execution time per iteration by about 6 times.

serious adverse effects on a patient. The human factor implications of this speed-up would be evaluated in training scenarios using simulated patients once the whole system is developed. Thus *EMSContExt* outperforms MetaMap in terms of both F1 score and execution time. Also, *EMSContExt* is a suitable lightweight solution for integration to embedded/wearable devices deployed in resource-constrained (e.g., limited battery life, no or low network connectivity), and safety-critical situations.

## 5 Related Work

**Lexicon Expansion:** It is a widely studied topic in knowledge representation and AI, particularly in the context of sentiment analysis and polarity detection (Hamilton et al. 2016). The approaches can be roughly categorized into two classes: (i) the knowledge-driven approaches where a lexicon is expanded using a structured knowledge base, i.e., a thesaurus or an ontology (Lindes, Lonsdale, and Embley 2015), and (ii) the data-driven approaches where a lexicon is expanded by recognizing patterns from a corpus (Hamilton et al. 2016; Qadir et al. 2015). Some existing works address the low training resources issue utilizing distant or weak supervision. One potential limitation of applying these approaches for the EMS domain is that, unlike sentiment and polarity lexicons, the EMS lexicon cannot be categorized into a few classes.

**Medical Concept Extraction:** Since the EMS domain has a significant overlap with the medical domain, we consider related works in medical concept extraction. Usually, the generic concept extraction techniques cannot be directly applied to medical text as it often contains ambiguity, polysemy, and variation of word orders. So, several existing research focuses on information extraction methods from the medical text. Authors in (Devarakonda and Tsou 2015) developed an SVM classifier using lexical, medical, and structural features to extract patient’s clinical problems from electronic medical records.

Another relevant research area is medical concept mapping, i.e., mapping the mention of a variable-length medical concept to a standard term or to a medical concept in an external coding system. (Limsopatham and Collier 2016) focused on normalizing adverse drug reactions mentioned in social media to formal medical concepts using word embedding based features. Other approaches for medical concept normalization include proximity matching (McCallum, Bellare, and Pereira 2012), heuristic string matching rules

(Aronson and Lang 2010), and machine learning and deep learning based approaches (Luo et al. 2018) to normalize specific types of clinical concepts, such as disease, diagnostic and treatment procedure, etc.

Since the EMS data has some unique characteristics (e.g., domain-specific terms, low training resources), the aforementioned research can not be directly applied for EMS concept extraction. Also, these approaches often rely on classifiers with only a few class of concepts with a reasonable amount of annotated data for each class. However, the target set of EMS concepts can not be unambiguously categorized in a few classes, and the annotation is very expensive.

## 6 Conclusion

*EMSContExt* automatically extracts safety-critical EMS concepts from the first responders' narrations to enable real-time cognitive assistance for emergency response. Our extensive evaluation using a corpus of over 9400 real EMS narrations demonstrated the contribution of our weakly-supervised knowledge-integrated and data-driven lexicon expansion techniques to the performance of *EMSContExt*. The resulting solution is a standalone EMS protocol-specific concept extraction method that addresses the challenges of domain mismatch and low training resources. It can be easily integrated into a device in resource-constrained settings (i.e., wearable devices running concept extraction offline). Also, our solution outperforms the state-of-the-art medical concept extraction technique and only requires one-time, minimal supervision from domain experts. It also shows promising results for providing decision support through protocol-specific intervention suggestions using protocol selection and execution. In the future, *EMSContExt* will be evaluated using speech data collected from realistic simulations of emergency incidents to measure its noise tolerance, and human factor implications. While *EMSContExt* is developed for the EMS domain, the underlying techniques can be adapted to develop intelligent applications for other low-resource, and cognitively demanding domains.

## 7 Acknowledgments

This work was supported by the award 60NANB17D162 from the U.S. Department of Commerce, National Institute of Standards and Technology (NIST).

## References

Aronson, A. R., and Lang, F.-M. 2010. An overview of metemap: historical perspective and recent advances. *Journal of the American Medical Informatics Association* 17(3):229–236.

Church, K. W., and Hanks, P. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics* 16(1):22–29.

Devarakonda, M., and Tsou, C.-H. 2015. Automated problem list generation from electronic medical records in ibm watson. In *IAAI 2015*.

Duong, H. V.; Herrera, L. N.; Moore, J. X.; Donnelly, J.; Jacobson, K. E.; Carlson, J. N.; Mann, N. C.; and Wang, H. E. 2018. National characteristics of emergency medical services responses for older adults in the united states. *Prehospital emergency care* 22(1):7–14.

Hamilton, W. L.; Clark, K.; Leskovec, J.; and Jurafsky, D. 2016. Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Johnson, A. E.; Pollard, T. J.; Shen, L.; Li-wei, H. L.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L. A.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3:160035.

Limsopatham, N., and Collier, N. 2016. Normalising medical concepts in social media texts by learning semantic representation. In *2016 ACL*.

Lindes, P.; Lonsdale, D. W.; and Embley, D. W. 2015. Ontology-based information extraction with a cognitive agent. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Luo, Y.; Song, G.; Li, P.; and Qi, Z. 2018. Multi-task medical concept normalization using multi-view convolutional neural network. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

McCallum, A.; Bellare, K.; and Pereira, F. 2012. A conditional random field for discriminatively-trained finite-state string edit distance. *arXiv preprint arXiv:1207.1406*.

Mehrabi, S.; Krishnan, A.; Sohn, S.; Roch; et al. 2015. Deepen: A negation detection system for clinical text incorporating dependency relation into negex. *Journal of biomedical informatics* 54:213–219.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Miller, G. A. 1995. Wordnet: a lexical database for english. *Communications of the ACM* 38(11):39–41.

Mujjiga, S.; Krishna, V.; Chakravarthi, K.; and Vijayananda, J. 2019. Identifying semantics in clinical reports using neural machine translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33.

Pennington, J.; Socher, R.; and Manning, C. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 EMNLP*.

Preum, S. M.; Shu, S.; Ting, J.; Lin, V.; Williams, R.; Stankovic, J.; and Alemzadeh, H. 2018. Towards a cognitive assistant system for emergency response. In *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*, 347–348.

Preum, S.; Shu, S.; Hotaki, M.; Williams, R.; Stankovic, J.; and Alemzadeh, H. 2019. CognitiveEMS: A cognitive assistant system for emergency medical services. *ACM SIGBED Review* 16(2):51–60.

Qadir, A.; Mendes, P. N.; Gruhl, D.; and Lewis, N. 2015. Semantic lexicon induction from twitter with pattern relatedness and flexible term length. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Shu, S.; Preum, S.; Pitchford, H.; Williams, R.; Stankovic, J.; and Alemzadeh, H. 2019. A behavior tree cognitive assistant system for emergency medical services. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Soysal, E.; Wang, J.; Jiang, M.; Wu, Y.; Pakhomov, S.; Liu, H.; and Xu, H. 2017. Clamp—a toolkit for efficiently building customized clinical natural language processing pipelines. *Journal of the American Medical Informatics Association* 25(3):331–336.

Uzuner, Ö.; South, B. R.; Shen, S.; and DuVall, S. L. 2011. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association* 18(5):552–556.